

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université A/MIRA de Bejaia
Faculté des Sciences de la Nature et de la Vie
Département TCSN

Chapitre 1 : Statistiques Descriptives

Pour les étudiants L2 en Biologie

Présenté par : **Dr. Hacene GHAROUT**

Table des matières

Table des matières	1
1 Statistiques Descriptives	3
1.1 Introduction	3
1.1.1 Concepts de base	3
1.1.2 Types de caractères	4
1.2 Tableaux statistiques et représentations graphiques	4
1.2.1 Tableau statistique relatif à un caractère qualitatif et sa représentation graphique	5
1.2.2 Tableaux statistiques relatifs à un caractère quantitatif et représentations graphiques	7
1.3 Fréquences relatives cumulées et effectifs cumulés	11
1.3.1 Variable statistique discrète	11
1.3.2 Variable statistique continue	12
1.4 Paramètres d'une v.s.	13
1.4.1 Moyenne arithmétique	13
1.4.2 La médiane	15
1.4.3 Le mode	17
1.5 Paramètres de dispersion	18
1.5.1 Étendu	18
1.5.2 Variance et Écart type	18
1.5.3 Les quartiles	19
2 Analyse Combinatoire	23
2.1 Introduction	23
2.1.1 Notion de répétition	23
2.1.2 Notion d'ordre	23
2.1.3 Factoriel d'un nombre n	24

2.2	Arrangements	24
2.2.1	Arrangement sans répétition	24
2.2.2	Arrangement avec répétition	25
2.3	Permutations	25
2.3.1	Permutation Sans répétition	25
2.3.2	Permutation avec répétition	26
2.4	Combinaisons	26
2.4.1	Combinaison sans répétitions (sans remises)	26
2.4.2	Combinaison avec répétitions (avec remises)	27
2.4.3	Propriétés des Combinaisons et Binôme de Newton	27

Chapitre 1

Statistiques Descriptives

1.1 Introduction

La statistique descriptive est un ensemble de méthodes permettant de décrire et d'analyser des phénomènes susceptibles d'être dénombrés et classés. Elle a pour but de décrire et non d'expliquer.

1.1.1 Concepts de base

Les observations constituent la source des informations statistiques. Avant de débiter l'étude, il faut définir l'ensemble étudié et les critères de la description chiffrée.

1. Les ensembles étudiés sont appelés **population**.
2. Les éléments de la population sont appelés **individus ou unités statistiques**.
3. Les sous-ensembles de la population est **un échantillon** et sa taille correspond à son cardinal.
4. Les critères étudiés constituent des caractères ; et un caractère permet de déterminer une partition de la population.

Exemple 1.1. Nous résumons les différents concepts dans cet exemple :

Population : l'ensemble des tous les employés d'une usine.

Individu : chaque employé est un individu.

Caractère : le salaire, l'état matrimonial, le nombre d'enfants,... etc.

Les modalités du caractère : marié, célibataire, divorcé et veuf sont les modalités de l'état matrimonial.

1.1.2 Types de caractères

On distingue deux types de caractères : qualitatif et quantitatif.

Caractère qualitatif

Définition 1.1. Un caractère est dit qualitatif lorsque ses modalités ne sont pas mesurables.

Exemple 1.2. Les couleurs du pelage : l'ensemble des modalités est

$$\{\text{noir, marron, blanc, } \dots\}.$$

Caractère quantitatif

Définition 1.2. Un caractère est dit quantitatif lorsque ses modalités sont des nombres. On lui donne souvent le nom de **variable statistique** (v.s.).

Une v.s. peut être :

Discrète : Si elle prend des valeurs isolées.

Exemple 1.3. Le nombre d'enfants d'une famille.

Continue : Lorsque elle peut prendre n'importe quelle valeur dans son domaine de variation.

Remarque 1.1. Dans le cas continu, le nombre de ces valeurs est toujours très, de ce cas, on regroupe toutes ces valeurs en classes.

En général, toutes les grandeurs liées à l'espace (longueur, surface, volume, ...), au temps (age), à la masse (poids, teneur, ...) ou à des combinaisons (vitesse, débit, ...) sont des variables statistiques continues.

1.2 Tableaux statistiques et représentations graphiques

Soit une population composée de n individus, sur laquelle on a étudié un caractère possédant k valeurs possibles. Ces valeurs x_1, x_2, \dots, x_k sont des modalités (cas qualitatif) ou des nombres (cas quantitatif).

Soient :

n_1 le nombre d'individus ayant pris la valeur x_1
 n_2 le nombre d'individus ayant pris la valeur x_2
 \vdots
 n_k le nombre d'individus ayant pris la valeur x_k

n_i est appelé **fréquence** ou **effectif** de la valeur x_i et n est l'effectif total.

On appelle **fréquence relative** ou **effectif relatif** de la valeur x_i la quantité :

$$f_i = \frac{n_i}{n} \text{ (ou en \% } f_i = \frac{n_i}{n} \times 100\%).$$

C'est la proportion d'individus ayant pris la valeur x_i .

Remarque 1.2.

$$\sum_{i=1}^k n_i = n_1 + n_2 + \dots + n_k = n.$$

$$\sum_{i=1}^k f_i = f_1 + f_2 + \dots + f_k = \frac{n_1}{n} + \frac{n_2}{n} + \dots + \frac{n_k}{n} = 1.$$

Modalités	Effectif	Fréquences relatives
x_i	n_i	$f_i = \frac{n_i}{n}$
x_1	n_1	$f_1 = \frac{n_1}{n}$
x_2	n_2	$f_2 = \frac{n_2}{n}$
\vdots	\vdots	\vdots
x_k	n_k	$f_k = \frac{n_k}{n}$
Total	$\sum_{i=1}^k n_i = n$	$\sum_{i=1}^k f_i = 1$

Tableau des effectifs et des fréquences relatives

1.2.1 Tableau statistique relatif à un caractère qualitatif et sa représentation graphique

Exemple 1.4. On veut étudier les lois de Mendel sur le caractère couleur de la fleur de Balsamine. Pour cela on étudiera le croisement des plantes hétérozygotes. On obtient 4 couleurs (modalités x_1, x_2, x_3 et x_4).

Population : les plantes de Balsamine.

Individu : une plante.

Caractère étudié : couleur de la fleur.

Modalités x_i	Effectif n_i	Fréquences relatives f_i	f_i en %
Pourpre	1790	0.5778	57.78 %
Rose	547	0.1766	17.66 %
Blanc-Lavande	548	0.1769	17.69 %
Blanche	213	0.0688	6.88%
Total	3098	1	100 %

Représentation graphique

L'information résumée dans un tableau statistique se traduit par un graphique pour en réaliser une synthèse visuelle.

a) Représentation par tuyaux d'orgue (diagramme en colonnes)

Dans ce cas, le graphe s'obtient en construisant autant de colonnes que des modalités du caractère qualitatif. Ces colonnes sont des rectangles de bases constantes et de hauteurs proportionnelles aux fréquences relatives.

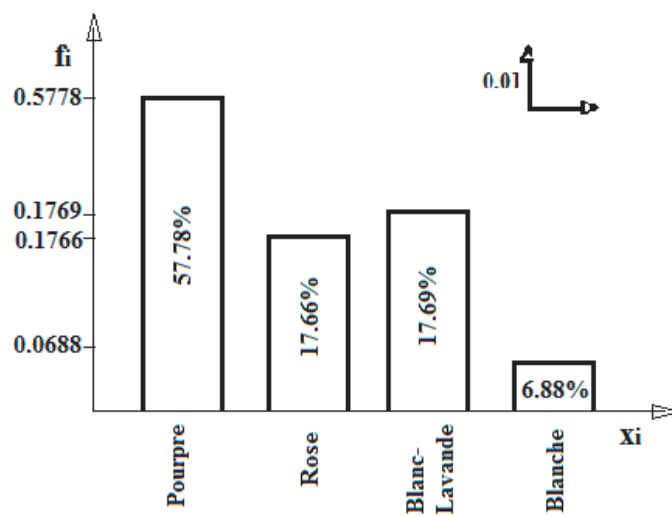


FIGURE 1.1 – Représentation en tuyaux d'orgue des fréquences relatives.

b) Représentation par le diagramme circulaire (camembert)

$$\theta_i = 360^\circ \times f_i = 360^\circ \times \frac{n_i}{n}.$$

Les angles correspondant de l'exemple sont :

$$\theta_1 = 360^\circ \times 0.5778 = 208.01^\circ \longrightarrow \text{Pourpre};$$

$$\theta_2 = 360^\circ \times 0.1766 = 63.58^\circ \longrightarrow \text{Rose};$$

$$\theta_3 = 360^\circ \times 0.1769 = 63.68^\circ \longrightarrow \text{Blanc-Lavande};$$

$$\theta_4 = 360^\circ \times 0.0688 = 24.77^\circ \longrightarrow \text{Blanche.}$$

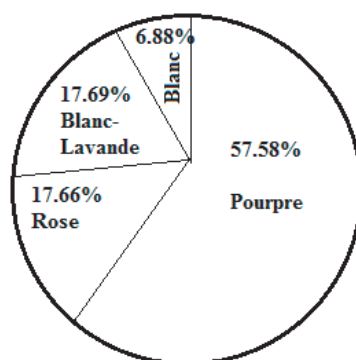


FIGURE 1.2 – Diagramme en camembert des fréquences relatives.

1.2.2 Tableaux statistiques relatifs à un caractère quantitatif et représentations graphiques

1) Cas d'une variable statistique discrète :

Exemple 1.5. Lors d'un contrôle d'une chaîne de médicaments, on s'intéresse au nombre de comprimés défectueux dans un lot. L'étude de 200 lots a donné les résultats suivants :

75 lots ont 0 comprimés défectueux ;

53 lots ont 1 comprimé défectueux ;

39 lots ont 2 comprimés défectueux ;

23 lots ont 3 comprimés défectueux ;

9 lots ont 4 comprimés défectueux ;

1 lot a 5 comprimés défectueux.

Population : l'ensemble des lots des médicaments.

Individu : un lot.

Caractère étudié : nombre de comprimés défectueux.

Modalités : 0, 1, 2, 3, 4 et 5.

Les fréquences relatives obtenues sont données dans le tableau suivant :

Modalités (Nbre de comprimés défectueux)	Nbre de lots n_i	Fréq. rel. $f_i = \frac{n_i}{n}$	Fréq. rel. en %
0	75	0.375	37.5 %
1	53	0.265	26.5 %
2	39	0.195	19.5 %
3	23	0.115	11.5 %
4	9	0.045	4.5 %
5	1	0.005	0.5 %
Total	200	1	100%

Représentation graphique :

On utilise le diagramme en batons pour représenter les effectifs n_i et les fréquences relatives f_i . Dans le cas du graphe des fréquences relatives, en joignant les sommets des batons on obtient le polygone des fréquences relatives.

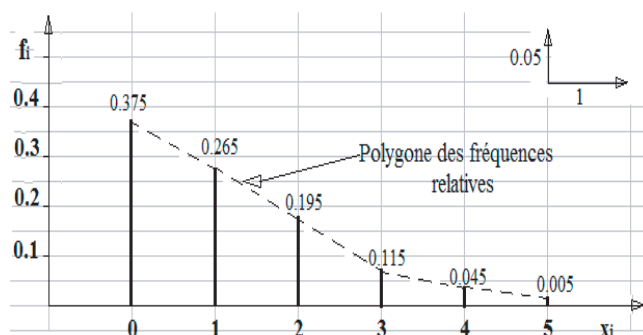


FIGURE 1.3 – Diagramme en escalier des fréquences relatives.

2) Cas d'une variable statistique continue :

Lorsque la v.s. est continue les données sont regroupées en classes

$$[e_0, e_1[, [e_1, e_2[, [e_2, e_3[, \dots, [e_{k-1}, e_k[.$$

Les modalités x_i représentent les centres c_i des classes $[e_{i-1}, e_i[$, avec :

- $c_i = \frac{e_{i-1} + e_i}{2}$, $i \in \{1, 2, \dots, k\}$;
- e_{i-1} est appelé l'extrémité inférieure de la classe $[e_{i-1}, e_i[$;
- e_i est appelé l'extrémité supérieure de la classe $[e_{i-1}, e_i[$;
- $a_i = e_i - e_{i-1}$ est l'amplitude de la classe $[e_{i-1}, e_i[$;
- Etendu = $e_k - e_0$.

Exemple 1.6. Une étude faite sur la taille d'un groupe d'étudiants (en mètre) a donné les résultats suivants :

Classes	Centre c_i	Effectif n_i	Fréq. rel. f_i (f_i en %)
[1.5; 1.6[1.55	8	0.08 (8%)
[1.6; 1.7[1.65	33	0.33 (30%)
[1.7; 1.8[1.75	31	0.31 (31%)
[1.8; 1.9[1.85	22	0.22 (22%)
[1.9; 2[1.95	6	0.06 (6%)
Total		100	1 (100%)

Population : les étudiants du groupe.

Individu : un étudiant.

Caractère étudié : la taille d'un étudiant.

Représentation graphique :

Dans le cas d'une variable statistique continue on utilise l'histogramme.

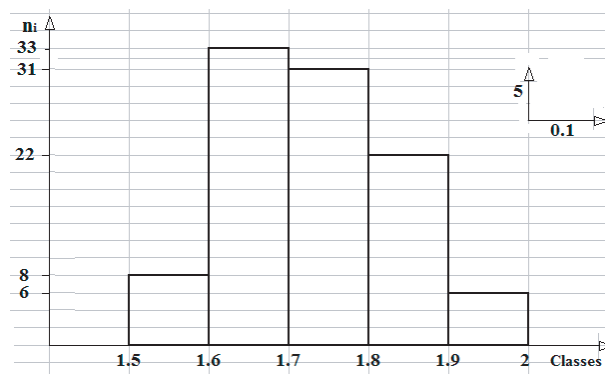


FIGURE 1.4 – L'histogramme des effectifs de l'exemple 1.6.

L'histogramme dans le cas des amplitudes inégales :

Dans ce cas les classes ont des amplitudes différentes. Du coup, il faut effectuer des corrections pour tenir compte des différences d'amplitude. Il convient de diviser les fréquences par leurs amplitudes correspondantes et on obtient, ainsi **l'amplitude corrigée** (h_i).

Exemple 1.7. Supposons que l'on regroupe les données de l'exemple précédent en classe d'amplitudes inégales.

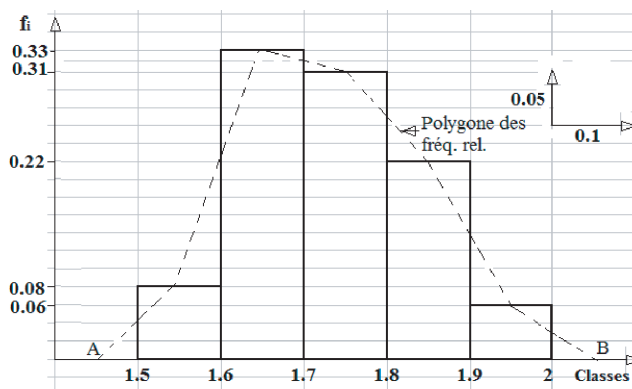


FIGURE 1.5 – L’histogramme des fréquences relatives de l’exemple 1.6.

Classes	Amplitude de la classe a_i	Effectif n_i	Fréq. rel. f_i	Amplitude corrigée $h_i = \frac{f_i}{a_i}$
[1.5; 1.7[0.2	41	0.41	2.05
[1.7; 1.8[0.1	31	0.31	3.1
[1.8; 2[0.2	28	0.28	1.4

Représentation graphique :

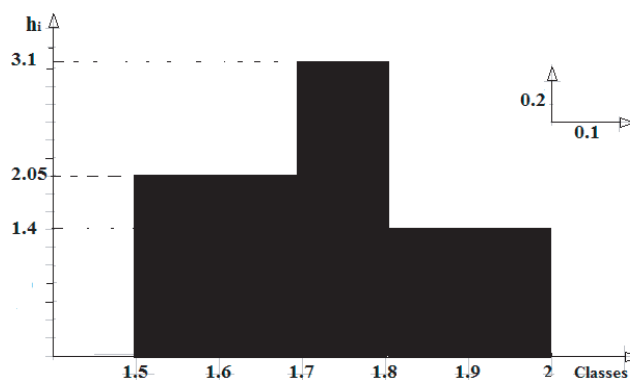


FIGURE 1.6 – Histogramme avec amplitudes inégales.

1.3 Fréquences relatives cumulées et effectifs cumulés

La fréquence relative cumulée F_i est la somme des fréquences relatives correspondantes aux valeurs de la v.s. inférieure à x_{i+1} .

$$\begin{aligned} F_1 &= f_1; \\ F_2 &= f_1 + f_2; \\ F_3 &= f_1 + f_2 + f_3; \\ &\vdots \\ F_i &= f_1 + f_2 + \dots + f_j + \dots + f_i. \end{aligned}$$

La fréquence relative cumulée F_i indique la proportion des individus pour lesquels la v.s. est inférieure à x_{i+1} .

De même façon on définit les effectifs cumulés

$$N_i = \sum_{j=1}^i n_j.$$

D'une manière équivalente, la fréquence relative cumulée $F_i = \frac{N_i}{n}$.

1.3.1 Variable statistique discrète

Exemple 1.8. On reprend l'exemple du cas discret (l'exemple 1.5 des comprimés défectueux).

moins de x_i	N_i	F_i
moins de 0	0	0
moins de 1	$0 + 75 = 75$	$0 + f_1 = 0.375$
moins de 2	$75 + 53 = 128$	$f_1 + f_2 = 0.64$
moins de 3	167	0.835
moins de 4	190	0.95
moins de 5	199	0.995
moins de x_k ($x_k > 5$)	$199 + 1 = 200$	1

Remarque 1.3.

$$\begin{aligned} \sum_{i=1}^k f_i &= 1; \\ F_i &= 0 \quad \text{si } x_i < x_1; \\ F_i &= 1 \quad \text{si } x_i > x_k, \end{aligned}$$

où x_1 est la plus petite valeur observée et x_k est la plus grande valeur observée.

Soit X une v.s. discrète et x_1, x_2, \dots, x_k les valeurs rangées dans l'ordre. La fonction de répartition d'une v.s. discrète est définie de \mathbb{R} dans $[0, 1]$ et est donnée par :

$$F(x) = \begin{cases} 0, & \text{si } x < x_1; \\ f_1, & \text{si } x_1 < x < x_2; \\ f_1 + f_2, & \text{si } x_2 < x < x_3; \\ \vdots & \\ f_1 + f_2 + \dots + f_i, & \text{si } x_i < x < x_{i+1}; \\ \vdots & \\ 1, & \text{si } x \geq x_k. \end{cases}$$

Exemple 1.9. Écrivons la fonction de répartition de la v.s. X de l'exemple des comprimés défectueux (l'exemple 1.5) :

$$F(x) = \begin{cases} 0, & \text{si } x < 0; \\ 0.375, & \text{si } 0 < x < 1; \\ 0.64, & \text{si } 1 < x < 2; \\ 0.835, & \text{si } 2 < x < 3; \\ 0.95, & \text{si } 3 < x < 4; \\ 0.995, & \text{si } 4 < x < 5; \\ 1, & \text{si } x \geq 5. \end{cases}$$

La courbe cumulative est la représentation graphique des fréquences relatives cumulées. Dans le cas discret, la courbe cumulative est une courbe en escalier (voir figure 1.7), dont les paliers horizontaux ont pour coordonnées (x_i, F_i) .

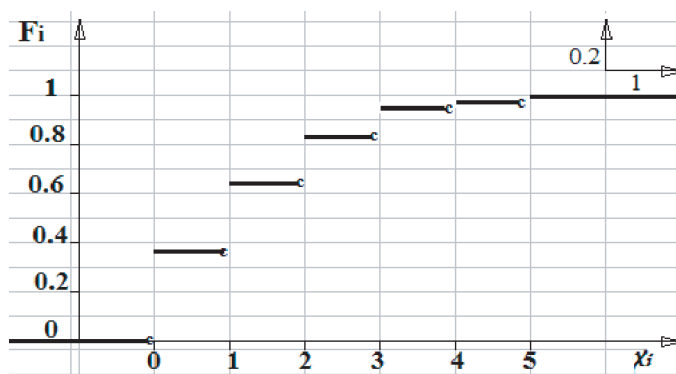


FIGURE 1.7 – Courbe des fréquences relatives cumulées (courbe cumulative).

1.3.2 Variable statistique continue

Exemple 1.10. On reprend l'exemple précédent sur la taille des étudiants. Les effectifs et les fréquences relatives cumulées sont données dans le tableau suivant :

moins de x_i	Effectifs cumulés N_i	Fréq. relatives $F_i = \frac{N_i}{n}$	cumulées (en %)
moins de 1.5	0	0	(0 %)
moins de 1.6	8	0.08	(8 %)
moins de 1.7	41	0.41	(41 %)
moins de 1.8	72	0.72	(72 %)
moins de 1.9	94	0.94	(94 %)
moins de x ($x \geq 2$)	100	1	(100 %)

La courbe cumulative est donnée dans la figure suivante :

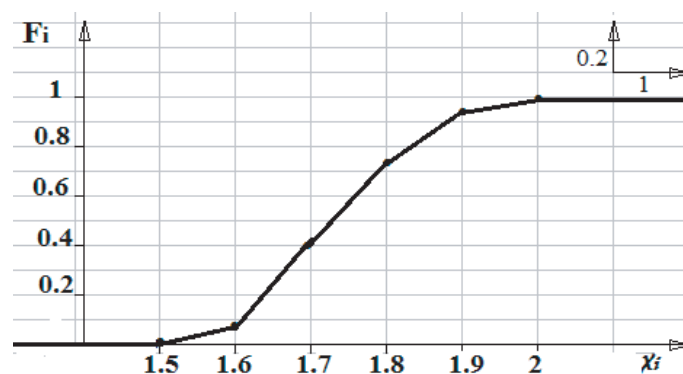


FIGURE 1.8 – Courbe des fréquences relatives cumulées du cas continu.

1.4 Paramètres d'une v.s.

Lorsque on observe une représentation graphique d'une série statistique, on peut en tirer deux observations :

1. L'ordre de grandeur de la v.s. caractérisée par des valeurs situées au centre de la distribution, se sont les paramètres de tendance centrale ou de position.
2. Les fluctuations des observations autour de la valeur centrale, mesurées par des écarts à celles-ci, se sont les paramètres de dispersion.

1.4.1 Moyenne arithmétique

La moyenne arithmétique est la somme de toutes les valeurs observées divisée par le nombre total des observations.

(a) Cas d'une v.s. discrète (données non groupées) :

Soit X une v.s. discrète et x_1, x_2, \dots, x_k ses valeurs, auquel correspondent les effectifs

n_1, n_2, \dots, n_k ; avec $n = \sum_{i=1}^k n_i$ l'effectif total.

La moyenne arithmétique notée \bar{x} de cette série stochastique, est définie par :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i.$$

Remarque 1.4.

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^k n_i x_i = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{n} \\ &= \frac{n_1}{n} x_1 + \frac{n_2}{n} x_2 + \dots + \frac{n_k}{n} x_k \\ &= f_1 x_1 + f_2 x_2 + \dots + f_k x_k \\ &= \sum_{i=1}^k f_i x_i, \end{aligned}$$

où f_i est la fréquence relative.

Exemple 1.11. La moyenne arithmétique de 200 lots de comprimés de l'exemple 1.5 est :

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^k n_i x_i \\ &= \frac{75.0 + 53.1 + 39.2 + 23.3 + 9.4 + 1.5}{200} = \frac{241}{200} = 1.205. \end{aligned}$$

(b) Cas d'une v.s. continue (données groupées) :

Dans le cas d'une v.s. continue, les observations sont groupées dans des classes; et nous avons la même formule que le cas discret, sauf qu'on remplace les x_i par les centres des classes x_i :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i c_i = \sum_{i=1}^k f_i c_i.$$

Exemple 1.12. La moyenne arithmétique de la taille des étudiants de l'exemple 1.6 est :

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^k n_i c_i \\ &= \frac{(8 \times 1.55) + (33 \times 1.65) + (31 \times 1.75) + (22 \times 1.85) + (6 \times 1.95)}{100} \\ &= \frac{173.5}{100} = 1.735 \text{ mètre.} \end{aligned}$$

(c) Propriétés de la moyenne arithmétique :

La moyenne de la somme des écarts d'un ensemble d'observations à leurs moyenne arithmétique est nulle $\frac{1}{n} \sum_{i=1}^k n_i(x_i - \bar{x}) = 0$.

En effet,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^k n_i(x_i - \bar{x}) &= \frac{1}{n} \sum_{i=1}^k (n_i x_i - n_i \bar{x}) = \frac{1}{n} \sum_{i=1}^k n_i x_i - \frac{1}{n} \sum_{i=1}^k n_i \bar{x} \\ &= \frac{1}{n} \sum_{i=1}^k n_i x_i - \bar{x} \frac{1}{n} \sum_{i=1}^k n_i = \frac{1}{n} \sum_{i=1}^k n_i x_i - \bar{x} \frac{n}{n} \\ &= \frac{1}{n} \sum_{i=1}^k n_i x_i - \bar{x} = \bar{x} - \bar{x} = 0. \end{aligned}$$

1.4.2 La médiane

La médiane est la valeur pour laquelle y a le même nombre d'individus à gauche et à droite dans un échantillon. Elle correspond au milieu de la distribution.

(a) Cas d'une variable statistique discrète :

Pour déterminer la médiane d'un échantillon ou d'une population :

1. on classe les individus par ordre croissant ;
2. on prend celui du milieu.

Remarque 1.5. la médiane Me est la valeur qui se trouve au centre de la série statistique :

- Si la valeur de l'effectif total n est impaire, i.e. $n = 2p + 1$, alors la médiane Me est la valeur qui se trouve à l'ordre $p + 1$ (ou $\frac{n+1}{2}$) :

$$Me = x_{p+1} = x_{\frac{n+1}{2}}.$$

- Si la valeur de l'effectif total n est paire, i.e. $n = 2p$, alors la médiane Me est la moyenne des valeurs qui se trouve à l'ordre p et $p + 1$ (ou $\frac{n}{2}$ et $\frac{n}{2} + 1$) :

$$Me = \frac{x_p + x_{p+1}}{2} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}.$$

Exemple 1.13. Soit un échantillon de 11 personnes dont le poids en kg est :

45, 68, 89, 74, 55, 62, 56, 74, 49, 52, 63.

Les poids classés par ordre croissant sont :

$$\underbrace{45, 68, 49, 52, 55, 56}_5, \underbrace{62}_{x_6=Me}, \underbrace{63, 68, 74, 74, 89}_5.$$

Si le nombre d'individus est pair, $n = 12$, on prend la moyenne entre les deux valeurs centrales.

$$\underbrace{45, 68, 49, 52, 55, 55, 56}_6, \underbrace{62, 63, 68, 74, 74, 89}_6.$$

$$45, 68, 49, 52, 55, 55, \underbrace{56, 62}_{Me}, 63, 68, 74, 74, 89.$$

$$\text{La médiane } Me = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{x_{\frac{12}{2}} + x_{\frac{12}{2}+1}}{2} = \frac{x_6 + x_7}{2} = \frac{56 + 62}{2} = 59 \text{ kg.}$$

(b) Cas d'une v.s. continue :

Dans ce cas la médiane est donnée par la formule suivante :

$$Me = L_i + \left(\frac{\frac{n}{2} - \sum_{i=1}^{<Me} n_i}{n_{Me}} \right) \cdot a$$

- L_i : borne inférieure de la classe médiane (classe qui divise l'effectif en deux) ;
- n : effectif total ;
- $\sum_{i=1}^{<Me} n_i$: somme des effectifs correspondant à toutes les classes inférieures à la classe médiane ;
- n_{Me} : effectif de la classe médiane ;
- a : amplitude de la classe médiane.

Exemple 1.14. Prenons le tableau des fréquences relatives cumulées de l'exemple 1.6. Pour déterminer la classe médiane, il suffit de tirer une ligne horizontale partant du point 0.5 (50%) de l'axe des fréquences relatives cumulées dans la courbe cumulative, arriver à l'ogive on descend une ligne verticale jusqu'à l'axe des x , et la classe où se situe le point d'intersection est la classe médiane. La classe médiane correspond, aussi, à la classe où les effectifs cumulés atteint ou dépasse pour la première fois le 50%.

Dans notre exemple, la classe médiane est la classe [1.7; 1.8[.

- $L_i = 1.7$;
- $n = 100$;
- $\sum_{i=1}^{<Me} n_i = 33 + 8$;
- $n_{Me} = 31$;
- $a = 0.1$;

$$Me = L_i + \left(\frac{\frac{n}{2} - \sum_{i=1}^{<Me} n_i}{n_{Me}} \right) \cdot a = 1.7 + \left(\frac{\frac{100}{2} - (33 + 8)}{31} \right) 0.1 = 1.7290 \text{ mètre.}$$

1.4.3 Le mode

(a) Cas d'une v.s. discrète :

Le mode Mo d'un ensemble d'observations est l'observation que l'on rencontre le plus fréquemment et il correspond à la modalité x_i ayant le plus grand effectif.

Remarque 1.6. Une distribution observée peut avoir plusieurs modes. Lorsqu'une distribution observée possède un seul mode, on parle de distribution unimodale. Lorsqu'une distribution observée possède deux modes, on parle de distribution bimodale.

Exemple 1.15. 1. Dans l'exemple des comprimés défectueux (exemple 1.5), le mode est 0.

2. Dans l'exemple des observations suivantes : 2, 2, 3, 5, 6, 6, 7, 8, 9, 7, 10 ; y a trois modes : 2, 6 et 7.

(b) Cas d'une v.s. continue :

Dans le cas des données groupées en classes, le mode se calcule par la formule :

$$Mo = L_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) \cdot a$$

- L_i : borne inférieure de la classe modale (classe correspondant au plus grand effectif) ;
- Δ_1 : excédent de l'effectif de la classe modale par rapport à l'effectif de la classe précédente ;
- Δ_2 : excédent de l'effectif de la classe modale par rapport à l'effectif de la classe suivante ;
- a : amplitude de la classe modale.

Exemple 1.16. Prenons les données de l'exemple du cas quantitatif continu (exemple 1.6) et calculons le mode.

- La classe modale est $[1.6; 1.7[$;
- $L_i = 1.6$;
- $\Delta_1 = 33 - 8$;
- $\Delta_2 = 33 - 31$;
- $a = 0.1$.

Le mode est

$$Mo = L_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) \cdot a = 1.6 + \left(\frac{33 - 8}{(33 - 8) + (33 - 31)}\right) \times 0.1 = 1.6926 \text{ mètres.}$$

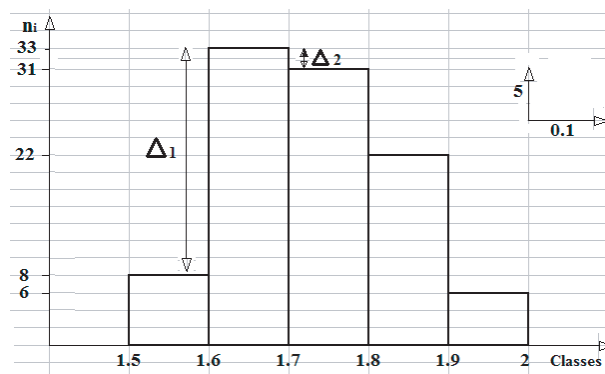


FIGURE 1.9 – L’histogramme des effectifs avec une illustration de Δ_1 et Δ_2 .

1.5 Paramètres de dispersion

Les deux ensembles d’observations suivants :

$$X = \{6, 6, 7, 7, \underbrace{8}_{\bar{x}=Me}, 9, 9, 10, 10\} \text{ et } Y = \{1, 2, 4, 6, \underbrace{8}_{\bar{y}=Me}, 10, 12, 14, 15\}$$

ont la même moyenne et la même médiane $\bar{x} = \bar{y} = Me = 8$, mais ils sont différents. Le premier ensemble est moins dispersé que le deuxième.

1.5.1 Étendu

On appelle étendu, notée e , la différence entre la plus grande valeur et la plus petite valeur observée.

Exemple 1.17. L’étendu de X est $e = 10 - 6 = 4$ et l’étendu de Y est $e = 15 - 1 = 14$.

1.5.2 Variance et Écart type

- **Variance**

La variance notée $V(X)$ est la moyenne des carrés des écarts des observations à la moyenne. Elle est définie par :

$$V(X) = \frac{1}{n} \sum_1^k n_i (x_i - \bar{x})^2 = \sum_1^k f_i (x_i - \bar{x})^2.$$

Propriétés

- 1) $V(X) \geq 0$;
- 2) $V(X) = \frac{1}{n} \sum_1^k n_i x_i^2 - \bar{x}^2 = \sum_1^k f_i x_i^2 - \bar{x}^2$.
- 3) Lorsque on compare les observations de deux variables statistiques X et Y , celle qui l'écart-type le plus élevé est la plus dispersée.

• Écart type

L'écart-type noté σ_X est la racine carrée de $V(X)$:

$$\sigma_X = \sqrt{V(X)} = \sqrt{\sum_1^k f_i (x_i - \bar{x})^2}.$$

Exemple 1.18. Cas quantitatif Discret

La variance des 200 lots de médicaments (exemple 1.5)) est :

$$\begin{aligned} V(X) &= \frac{1}{n} \sum_1^k n_i x_i^2 - \bar{x}^2 \\ &= \frac{(75 \times 0^2) + (53 \times 1^2) + (39 \times 2^2) + (23 \times 3^2) + (9 \times 4^2) + (1 \times 5^2)}{200} - (1.205)^2 \\ &= 1.473 \end{aligned}$$

et l'écart-type $\sigma_X = \sqrt{V(X)} = \sqrt{1.473} = 1.214$.

Exemple 1.19. Cas quantitatif Continu

La variance de l'exemple 1.6 est :

$$\begin{aligned} V(X) &= \frac{1}{n} \sum_1^k n_i c_i^2 - \bar{x}^2 \\ &= \frac{(8 \times 1.55^2) + (33 \times 1.65^2) + (31 \times 1.75^2) + (22 \times 1.85^2) + (6 \times 1.95^2)}{100} - (1.735)^2 \\ &= 0.010875. \end{aligned}$$

L'écart-type $\sigma_X = \sqrt{V(X)} = \sqrt{0.010875} = 0.1043$.

1.5.3 Les quartiles

On utilise couramment les quartiles Q_1 , Q_2 et Q_3 .

Q_1 est le quartile d'ordre $\frac{1}{4}$, représente 25% de l'échantillon ;

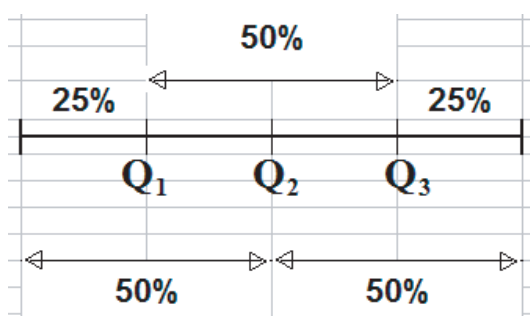
Q_2 est le quartile d'ordre $\frac{1}{2}$, représente 50% de l'échantillon ;

Q_3 est le quartile d'ordre $\frac{3}{4}$, représente 75% de l'échantillon.

Intervalle interquartile

(Q_1, Q_2) contient 50% de la population laissant à droite 25% et à gauche 25%. Cet intervalle est donnée par : $Q_1 - Q_2$.

Pour déterminer l'intervalle interquartile, il faut déterminer d'abord Q_1 et Q_3 .



Le premier quartile Q_1

(a) Cas discret

Q_1 est la valeur x_i dont le rang (la position) est le plus entier qui suit $\frac{n}{4}$.

Exemple 1.20. Dans l'exemple des observations suivantes : 2, 3, 4, 5, 6, 6, 7, 7, 8, 9, 10 ; on a :

$n = 11$ et $\frac{n}{4} = \frac{11}{4} = 2.25$. Le plus petit entier qui suit $\frac{n}{4} = 2.25$ est 3, alors Q_1 est la troisième valeur. D'où $Q_1 = x_3 = 4$.

Exemple 1.21. Reprenons l'exemple des comprimés défectueux (exemple 1.5), où on a $\frac{n}{4} = \frac{200}{4} = 50$ et la valeur pour laquelle les effectifs cumulés atteint et dépasse pour la première fois 50 est "moins 1", c'est à dire la valeur de 0, alors $Q_1 = 0$.

(b) Cas continu

Dans ce cas le premier quartile est donné par la formule suivante :

$$Q_1 = L_i + \left(\frac{\frac{n}{4} - \sum_{i=1}^{<Q_1} n_i}{n_{Q_1}} \right) \cdot a$$

- L_i : borne inférieure de la classe de Q_1 ;
- n : effectif total ;
- $\sum_{i=1}^{<Q_1} n_i$: somme des effectifs correspondant à toutes les classes inférieures à la classe de Q_1 ;

- n_{Q_1} : effectif de la classe de Q_1 ;
- a : amplitude de la classe de Q_1 .

Exemple 1.22. Prenons le tableau de l'exemple 1.6. Pour déterminer la classe de Q_1 , il suffit de tirer une ligne horizontale partant du point 0.25 (25%) de l'axe des fréquences relatives cumulées dans la courbe cumulative, arriver à l'ogive on descend une ligne verticale jusqu'à l'axe des x et la classe où se situe le point d'intersection est la classe de Q_1 .

La classe de Q_1 correspond, à la classe où les effectifs cumulés atteignent ou dépassent pour la première fois 25% de l'effectif total.

Dans notre exemple, la classe de Q_1 est la classe $[1.6; 1.7[$.

- $L_i = 1.6$;
- $n = 100$;
- $\sum_{i=1}^{<Q_1} n_i = 8$;
- $n_{Q_1} = 33$;
- $a = 0.1$;

$$Q_1 = L_i + \left(\frac{\frac{n}{4} - \sum_{i=1}^{<Q_1} n_i}{n_{Q_1}} \right) \cdot a = 1.6 + \left(\frac{\frac{100}{4} - 8}{33} \right) 0.1 = 1.65515 \text{ mètre.}$$

Le troisième quartile Q_3

(a) Cas discret

Q_3 est la valeur x_i dont le rang (la position) est le plus entier qui suit $\frac{3n}{4}$.

Exemple 1.23. Dans le même exemple des observations : 2, 3, 4, 5, 6, 6, 7, 7, 8, 9, 10 ; on a :

$n = 11$ et $\frac{3n}{4} = \frac{3 \times 11}{4} = 8.25$. Le plus petit entier qui suit $\frac{3n}{4} = 8.25$ est 9, alors Q_3 est la 9^{ème} valeur. D'où $Q_3 = x_9 = 8$.

Exemple 1.24. Dans l'exemple des comprimés défectueux (exemple 1.5), on a $\frac{3n}{4} = \frac{3 \times 200}{4} = 150$ et la valeur où les effectifs cumulés atteignent et dépassent pour la première fois 150 est "moins 3", c'est à dire la valeur de 2, alors $Q_3 = 2$.

(b) Cas continu

Dans le cas le troisième quartile est donné par la formule suivante :

$$Q_3 = L_i + \left(\frac{\frac{3n}{4} - \sum_{i=1}^{<Q_3} n_i}{n_{Q_3}} \right) \cdot a$$

- L_i : borne inférieure de la classe de Q_3 ;
- n : effectif total ;
- $\sum_{i=1}^{<Q_3} n_i$: somme des effectifs correspondant à toutes les classes inférieures à la classe de Q_3 ;
- n_{Q_3} : effectif de la classe de Q_3 ;
- a : amplitude de la classe de Q_3 .

Exemple 1.25. Pour déterminer la classe de Q_3 de l'exemple 1.6, également, il suffit de tirer une ligne horizontale partant du point 0.75 (75%) de l'axe des fréquences relatives cumulées dans la courbe cumulative, arriver à l'ogive on descend une ligne verticale jusqu'à l'axe des x et la classe où se situe le point d'intersection est la classe de Q_3 .

La classe de Q_3 correspond, à la classe où les effectifs cumulés atteignent ou dépassent pour la première fois 75% de l'effectif total.

Dans notre exemple, la classe de Q_3 est la classe [1.8; 1.9[.

- $L_i = 1.8$;
- $n = 100$;
- $\sum_{i=1}^{<Q_3} n_i = 8 + 33 + 31$;
- $n_{Q_3} = 22$;
- $a = 0.1$;

$$Q_3 = L_i + \left(\frac{\frac{3n}{4} - \sum_{i=1}^{<Q_3} n_i}{n_{Q_3}} \right) \cdot a = 1.8 + \left(\frac{\frac{3 \times 100}{4} - (8 + 33 + 31)}{22} \right) 0.1 = 1.8136 \text{ mètre.}$$

Chapitre 2

Analyse Combinatoire

L'analyse Combinatoire est une branche des mathématiques qui étudie comment compter les objets. Elle fournit des méthodes de dénombrements particulièrement utile en théorie des probabilités.

2.1 Introduction

L'analyse Combinatoire a pour but de compter les dispositions qui peuvent être formées à partir des éléments d'un ensemble fini d'objets. Un objet est caractérisé par :

- la place qu'il occupe dans la disposition ;
- le nombre de fois où il peut apparaître.

2.1.1 Notion de répétition

Si un élément apparaît plus d'une fois dans une disposition, on dit que la disposition est avec répétition ; sinon, la disposition est dite sans répétition.

2.1.2 Notion d'ordre

Une disposition est dite ordonnée, si lorsque à chaque fois qu'un élément change de place (ou de position) la disposition change.

Exemple 2.1. On considère un ensemble E ayant trois éléments $E = \{a, b, c\}$. Choisir deux éléments dans cet ensemble peut se faire de plusieurs façons différentes.

Le tableau suivant, nous donne tous les cas possibles :

disposition	avec répétition	sans répétition
avec ordre (ordonnée)	aa, ab, ac, ba, bb, bc, ca, cb, cc	ab, ac, ba, bc, ca, cb
sans ordre (non ordonnée)	aa, ab, ac, bb, bc, cc	ab, ac, bb, bc

2.1.3 Factoriel d'un nombre n

Étant donné un nombre n , le factoriel de n , noté $n!$ est :

$$n! = n \times (n - 1) \times (n - 2) \times \dots \times 3 \times 2 \times 1.$$

Par convention, on a : $0! = 1$.

Exemple 2.2.

$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120.$$

$$10! = 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 3628800.$$

2.2 Arrangements

Définition 2.1. Étant donné un ensemble E de n objets, un arrangement de p de ces objets est une suite **ordonnée** de p objets pris parmi ces n objets.

On distingue deux types d'arrangements : avec et sans répétition.

2.2.1 Arrangement sans répétition

On appelle arrangement sans répétition de n objets p à p , toute disposition **ordonnée** de p objets choisis parmi les n objets **sans répétitions**.

Nombre d'arrangements sans répétition :

Le nombre d'arrangements, noté A_n^p , est :

$$A_n^p = n \times (n - 1) \times (n - 2) \dots \times (n - p + 1) = \frac{n!}{(n - p)!},$$

avec $1 \leq p \leq n$.

Dans un arrangement sans répétition, les p objets de la liste sont tous distincts. Cela correspond à un tirage **sans remise** et **avec ordre**.

Exemple 2.3. Combien de mots de trois lettres ne contenant pas plus d'une fois la même lettre peut-on former avec les lettres de l'alphabet ?

$$A_{26}^3 = \frac{26!}{(26 - 3)!} = 26 \times 25 \times 24 = 15600 \text{ mots.}$$

2.2.2 Arrangement avec répétition

On appelle arrangement avec répétition de n objets p à p , toute disposition **ordonnée** de p objets choisi parmi les n objets **avec répétitions**.

Nombre d'arrangements avec répétition :

Le nombre d'arrangements est :

$$n = \underbrace{n \times n \times n \dots \times n}_{p \text{ fois}}$$

avec $1 \leq p \leq n$.

Dans un arrangement sans répétition, les p objets de la liste ne sont pas nécessairement tous distincts. Cela correspond à un tirage **avec remise** et **avec ordre**.

Exemple 2.4. Combien de mots de deux lettres peut-on former avec les lettres de l'alphabet ?

$$26^2 = 26 \times 26 = 676 \text{ mots.}$$

2.3 Permutations

Définition 2.2. Étant donné un ensemble E de n objets. On appelle permutation de n objets distincts **toute suite ordonnée** de n objets ou **tout arrangement** n à n de ces objets.

2.3.1 Permutation Sans répétition

C'est le cas particulier de l'arrangement sans répétition de p objets parmi n objets, lorsque $p = n$.

Le nombre de permutations de n objets est :

$$P_n = n!.$$

Remarque 2.1.

$$P_n = A_n^n = \frac{n!}{(n-n)!} = n!.$$

Exemple 2.5. Le nombre de manières de placer 8 convives (invités) autour d'une table est

$$P_8 = 8! = 40320 \text{ possibilités.}$$

2.3.2 Permutation avec répétition

Dans le cas où il existe k objets identiques parmi les n objets, alors

$$P_n = \frac{n!}{k!}.$$

Exemple 2.6. Le nombre de mots possibles (avec ou sans signification) que l'on peut formé en permutant les 8 lettres du mot "Quantité", est

$$P_8 = \frac{8!}{2!} = 20160 \text{ mots, car on a le t 2 fois.}$$

Et en considérant le mot "Natation", le nombre de mots possibles est

$$P_8 = \frac{8!}{2!2!2!} = 5040 \text{ mots, car on a le n 2 fois, le a 2 fois et le t 2 fois.}$$

2.4 Combinaisons

2.4.1 Combinaison sans répétitions (sans remises)

Définition 2.3. Étant donné un ensemble E de n objets. On appelle combinaisons de p objets **tout ensemble** de p objets pris parmi les n objets sans remise.

Le nombre de combinaisons de p objets parmi n et sans remise, est :

$$C_n^p = \frac{n!}{(n-p)!p!},$$

avec $1 \leq p \leq n$.

Remarque 2.2.

$$C_n^p = \frac{A_n^p}{p!} = \frac{n!}{(n-p)!p!}.$$

Exemple 2.7. Le tirage au hasard de 5 cartes dans un jeu de 32 cartes (main de poker) est une combinaison avec $p = 5$ et $n = 32$. Le nombre de tirages possibles est

$$C_{32}^5 = \frac{32!}{(32-5)!5!} = 409696.$$

Exemple 2.8. La formation d'une délégation de 2 étudiants parmi un un groupe de 20 constitue une combinaison avec $p = 2$ et $n = 20$. Le nombre de délégations possibles est

$$C_{20}^2 = \frac{20!}{(20-2)!2!} = 190.$$

2.4.2 Combinaison avec répétitions (avec remises)

Le nombre de combinaisons de p objets parmi n et avec remise, est :

$$C_{n+p-1}^p = \frac{(n+p-1)!}{p!(n-1)!}.$$

Exemple 2.9. Soit la constitution de mots de 3 lettres à partir d'un alphabet à 5 lettres avec remise.

Le nombre de mots est

$$C_{5+3-1}^3 = C_7^3.$$

On distingue 3 cas possibles :

- C_5^3 nombre de mots de 3 lettres différentes et sans ordre ;
- $2C_5^2$ nombre de mots de 2 lettres différentes et une lettre redondante ;
- C_5^1 nombre de mots de 3 lettres identiques ;

au total, on a $C_5^3 + 2C_5^2 + C_5^1 = C_7^3 = 35$ mots possibles.

2.4.3 Propriétés des Combinaisons et Binôme de Newton

(1)

$$C_n^0 = C_n^n = \frac{n!}{n!} = 1;$$

$\forall n \geq 1$, on a

$$C_n^1 = C_n^{n-1} = \frac{n!}{(n-1)!} = n;$$

$\forall n \geq 2$, on a

$$C_n^2 = C_n^{n-2} = \frac{n(n-1)}{2};$$

(2) Par récurrence, on déduit :

Si $0 \leq p \leq n$, on a

$$C_n^p = C_n^{n-p}.$$

Combinaisons composées ou Formule de Pascal

Si $0 \leq p \leq n-1$, on a

$$C_{n-1}^{p-1} + C_{n-1}^p = C_n^p.$$

(3) **Binôme de Newton**

Le théorème du Binôme de Newton donne l'expression générale du développement

de $(a + b)^n$.

$$\begin{aligned}(a + b)^n &= \sum_{p=0}^n C_n^p a^{n-p} b^p \\ &= C_n^0 a^n b^0 + C_n^1 a^{n-1} b^1 + \dots + C_n^n a^0 b^n \\ &= a^n + n a^{n-1} b + \dots + b^n.\end{aligned}$$

Exemple 2.10. Pour $n = 4$, on a

$$\begin{aligned}(a + b)^4 &= \sum_{p=0}^4 C_4^p a^{4-p} b^p \\ &= C_4^0 a^4 b^0 + C_4^1 a^3 b^1 + \dots + C_4^4 a^0 b^4 \\ &= a^4 + 4a^3 b + 6a^2 b^2 + 4ab^3 + b^4.\end{aligned}$$

Pour $n = 5$,

$$(a + b)^5 = a^5 + 5a^4 b + 10a^3 b^2 + 10a^2 b^3 + 5ab^4 + b^5.$$

Remarque 2.3. Lorsque $a = b = 1$, on a $(1 + 1)^n = 2^n$ et $2^n = \sum_{p=0}^n C_n^p$. Alors

$$2^n = C_n^0 + C_n^1 + C_n^2 + \dots + C_n^n.$$

Bibliographie

- [1] F. Carrat et A. Mallet. *Biostatistique*. Université Pierre et Marie Curie, 2013.
- [2] J.P. Lecoutre. *Statistique et probabilités - Cours et exercices corrigés*. DUNOD, 2012.
- [3] D. Meghlaoui. *Introduction à la Statistique Descriptive*. Ecole Préparatoire en Sciences Economiques Commerciales et des Sciences de Gestion de Constantine, 2010.
- [4] M. MERCIER. *Biostatistique et probabilités*. ELLIPSES, 2011.